Research Article

Uniqueness of the Level Two Bayesian Network Representing a Probability Distribution

Linda Smail

New York Institute of Technology, College of Arts and Sciences, P.O. Box 840878, Amman 11184, Jordan

Correspondence should be addressed to Linda Smail, lynda.smail@gmail.com

Bayesian Networks are graphic probabilistic models through which we can acquire, capitalize on, and exploit knowledge. they are becoming an important tool for research and applications in artificial intelligence and many other fields in the last decade. This paper presents Bayesian networks and discusses the inference problem in such models. It proposes a statement of the problem and the proposed method to compute probability distributions. It also uses D-separation for simplifying the computation of probabilities in Bayesian networks. Given a Bayesian network over a family I of random variables, this paper presents a result on the computation of the probability distribution of a subset S of I using separately a computation algorithm and D-separation properties. It also shows the uniqueness of the obtained result.

1. Introduction

Bayesian networks are graphical models for probabilistic relationships among a set of variables. They have been used in many fields due to their simplicity and soundness. They are used to model, represent, and explain a domain, and they allow us to update our beliefs about some variables when some other variables are observed, which is known as inference in Bayesian networks.

Given a Bayesian network [1] relative to the set $X_I = (X_i)_{i \in I}$ of random variables, we are interested in computing the joint probability distribution of a nonempty subset S (called target) of I.

The computation of the probability distribution of X_S requires marginalizing out a set of variables of $\overline{S} = I - S$ from the joint distribution P_I corresponding to the Bayesian network.

In large Bayesian networks, the computation of probability distributions and conditional probability distributions may require summations relative to very large subsets of I.

Consequently there is a need to order and segment, if possible, these computations into several computations that are less intensive and more accessible to a parallel treatment. These segmentations are related to the graphic properties of the Bayesian networks.

This paper describes the computation of P_S using a specific order described by a proposed algorithm, and the segmentations of P_S.

We consider discrete random variables, but the results presented here can be generalized to continuous random variables with the density of X_i relative to a finite measure μ_i (the summations will be replaced by integrations relative to those measures μ_i).

The paper is organized as follows. Section 2 introduces Bayesian networks and Level two Bayesian networks. We, then, present in Section 3 the inference problem in Bayesian networks and the proposed computation algorithm. Section 4 outlines D-Separation in Bayesian networks and describes graphical partitions that will allow the segmentations of the computations of probability distributions. Section 5 proves the uniqueness of the results obtained in Sections 3 and 4.

2. Bayesian Networks and level two Bayesian networks

2.1. Bayesian Networks

A Bayesian network (BN) is a family of random variables $(X_i)_{i \in I}$ such that:

(i) the set I is finite and endowed with a structure of a directed acyclic graph (DAG), G, where, for each i:

    (a) $p(i)$ is the set of parents of i ($p(i) = \{j; (j, i) \in G\}$)

    (b) $e(i)$ is the set of children of i ($e(i) = \{j; (i, j) \in G\}$)

    (c) $d(i)$ is the set of descendants of i ($d(i) = \{k; \exists (\ell_0, \dots, \ell_n)\ \ell_0 = i,\ \ell_n = k,\ \forall s \in \{1, \dots, n\}\ (\ell_{s-1}, \ell_s) \in G\}$);

(ii) for each i, X_i is independent of $(X_j)_{j \in I - [p(i) \cup d(i)]}$ conditional on $X_{p(i)}$ (for more details see, e.g., [1–5]).

We know that this is equivalent to the equality

$$P_I(x_I) = \prod_{i \in I} P_{i/p(i)}\left(x_i / x_{p(i)}\right),\tag{2.1}$$

where P_I is the joint probability distribution of $X_I = (X_i)_{i \in I}$ and $P_{i/p(i)}$ is the probability distribution of X_i conditional on $X_{p(i)} = (X_j)_{j \in p(i)}$.

The joint probability distribution corresponding to the BN in Figure 1 can be written as

$$P_I(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$$
$$= P_1(x_1)P_2(x_2)P_{3/1,2}(x_3/x_1, x_2)P_4(x_4)P_6(x_6) \tag{2.2}$$
$$\times P_{5/3,4}(x_5/x_3, x_4)P_{7/5,6}(x_7/x_5, x_6)P_{8/3,7}(x_8/x_3, x_7)P_{9/8}(x_9/x_8)P_{10/8}(x_{10}/x_8).$$
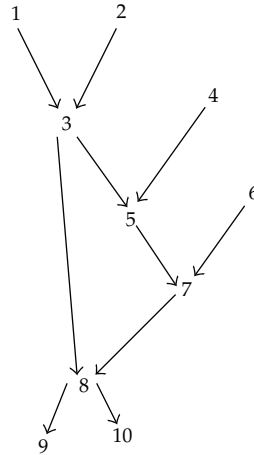
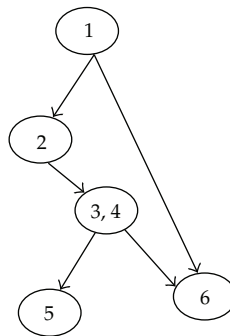**Figure 1:** An example of a Bayesian network.

Figure 2: Level two Bayesian network.

2.2. level two Bayesian networks

We consider the probability distribution P_I of a family $(X_i)_{i \in I}$ of random variables in a finite space $\Omega_I = \prod_{i \in I} \Omega_i$. Let $\mathcal{O}$ be a partition of I and let us consider a DAG $\mathcal{G}$ on $\mathcal{O}$.

We say that there is a link from J' to J'' (where J' and J'' are atoms of the partition $\mathcal{O}$) if $(J', J'') \in \mathcal{G}$. If $J \in \mathcal{O}$, we denote by $p(J)$ the set of parents of J, that is, the set of J' such that $(J', J) \in \mathcal{G}$.

The probability P_I is defined by the *Level Two Bayesian network* (BN2), on I, $(\mathcal{O}, \mathcal{G}, (P_{J/p(J)})_{J \in \mathcal{O}})$, if for each $J \in \mathcal{O}$, we have the conditional probability $P_{J/p(J)}$, or the probability of X_J conditioned by $X_{p(J)}$ (which, if $p(J) = \emptyset$, is the marginal probability P_J), so that

$$P_I(x_I) = \prod_{J \in \mathcal{O}} P_{J/p(J)}\left(x_J / x_{p(J)}\right). \tag{2.3}$$

The probability distribution P_I associated to the BN of level 2 in Figure 2 can be written as

$$P_I(x_I) = P_1(x_1)P_{2/1}(x_2/x_1)P_{3,4/2}(x_3,x_4/x_2)P_{5/3,4}(x_5/x_3,X_4)P_{6/1,3,4}(x_6/x_1,x_3,x_4). \quad (2.4)$$

2.3. Close Descendants

We define the set of close descendants of a node J (denoted $\mathrm{cd}(J)$) as the set of vertices containing the children of J and the vertices located on a path between J and one of its children.

In the example below (Figure 2), we have

$$\mathrm{cd}(1) = \{2,\{3,4\},6\} = \{2,3,4,6\}. \quad (2.5)$$

2.4. Initial Subset

For each subset S, we denote by S^+ the *initial subset* defined by S, that is, the set consisting of S itself and the S' such that there is a path in $\mathcal{G}$ from S' to S.

We can identify this subset with the union of all S' such that S' is an ancestor of S

For each S, the Initial subset S^+ is a BN, in other words the restriction of a BN to an initial subset is a BN

$$P_{S^+} = \prod_{s \in S^+} P_{s/p(s)}. \quad (2.6)$$

In the example above (Figure 2), we have

$$\{6\}^+ = \{1,2,\{3,4\}\} = \{1,2,3,4\}. \quad (2.7)$$

3. Inference in Bayesian Networks

Consider the BN in Figure 1.

Suppose we are interested in computing the distribution of $S = I - \{3\}$, in other words all variables are in the target S except X_3.

By marginalizing out the variables X_3 from the joint probability distribution P_I, the target distribution P_S can be written as

$$P_S(x_1,x_2,x_4,x_5,x_6,x_7,x_8,x_9,x_{10})$$

$$= P_1(x_1)P_2(x_2)P_4(x_4)P_6(x_6)P_{7/5,6}(x_7/x_5,x_6)P_{9/8}(x_9/x_8)$$

$$\times P_{10/8}(x_{10}/x_8)\sum_{x_3}[P_{3/1,2}(x_3/x_1,x_2)P_{5/3,4}(x_5/x_3,x_4)P_{8/3,7}(x_8/x_3,x_7)]$$

$$= P_1(x_1)P_2(x_2)P_4(x_4)P_6(x_6)P_{7/5,6}(x_7/x_5,x_6)P_{9/8}(x_9/x_8)P_{10/8}(x_{10}/x_8)\alpha(x_1,x_2,x_4,x_5,x_7).$$
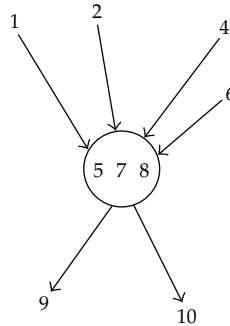
$$(3.1)$$

Figure 3: Level two Bayesian network on S.

α is a function that depends on X_1, X_2, X_4, X_5, and X_7 but has nothing to do with $P_{1,2,4,5,7}$ joint probability distribution of X_1, X_2, X_4, X_5, X_7.

By doing this we loose the structure of the BN.

If we do the marginalization as follow we obtain (according to Bayes' theorem):

$$P_S(x_1, x_2, x_4, x_5, x_6, x_7, X_8, X_9, X_{10})$$

$$= P_1(x_1)P_2(x_2)P_4(x_4)P_6(x_6)P_{9/8}(x_9/x_8)P_{10/8}(x_{10}/x_8)$$

$$\times \sum_{x_3}[P_{3/1,2}(x_3/x_1, x_2)P_{5/3,4}(x_5/x_3, x_4)P_{7/5,6}(x_7/x_5, x_6)P_{8/3,7}(x_8/x_3, x_7)]$$

$$= P_1(x_1)P_2(x_2)P_4(x_4)P_6(x_6)P_{9/8}(x_9/x_8)P_{10/8}(x_{10}/x_8)P_{5,7,8/1,2,4,6}(x_5, x_7, x_8/x_1, x_2, x_4, x_6).$$

$$(3.2)$$

In other words $\alpha(x_1, x_2, x_4, x_5, x_7)P_{7/5,6}(x_7/x_5, x_6) = P_{5,7,8/1,2,4,6}(x_5, x_7, x_8/x_1, x_2, x_4, x_6)$.

Which provides S with a structure of a level two Bayesian network shown in Figure 3.

The variables used in the marginalization above, to keep a structure of a BN2, is the the set of close descendants defined above.

More general, if we have to sum out more than one variable there is a need to order the variables first. The aim of the inference will be to find the marginalization, or elimination, ordering for the arbitrary set of variables not in the target. This aim is shared by other node elimination algorithms like "variables elimination" [6], "bucket elimination" [7].

The main idea of all these algorithms is to find the best way to sum over a set of variables from a list of factors one by one. An ordering of these variables is required as an input. The computation depends on the ordering elimination; different elimination ordering produce different factors.

The algorithm we proposed to solve this problem is called the "Successive Restrictions Algorithm" (SRA) [8]. SRA is a goal-oriented algorithm that tries to find an efficient marginalization (elimination) ordering for an arbitrary joint distribution.

The general idea of the algorithm of successive restrictions is to manage the succession of summations on all random variables out of the target S in order to keep on it a structure less constraining than the Bayesian network, but which allows saving in memory; that is
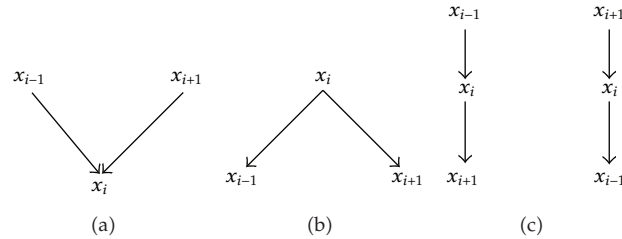
Figure 4: (a) A converging connexion. (b) A diverging connexion. (c) A serial connexion.

the structure of Bayesian Network of Level Two. This was possible using the notion of close descendants.

The principle of the algorithm was presented in details in [8].

4. D-Separation and Computations in a Bayesian Network

We have introduced an algorithm which makes possible the computation of the probability distribution of a subset of random variables $(X_s)_{s \in S}$ of the initial graph. It is also possible to use the SRA to compute any probability distribution of a set of variables X_A conditionally to another subset X_B $(P_{A|B})$.

This algorithm tries to achieve the target distribution by finding a marginalization ordering that takes into account the computational constraints of the application. It may happen that, in certain simple cases, the SRA would be less powerful than the traditional methods [6, 9–13], but it has the advantages of adapting to any subset of nodes of the initial graph, and also to present in each stage interpretable result in terms of conditional probabilities, and thus technically usable.

In addition to the SRA we propose, especially for large Bayesian networks, to segment the computations into several less heavier computations that could be carried independently. These segmentations are possible using the D-separation.

4.1. D-Separations and Classical Results

Consider a DAG (I, G). A chain is a sequence $(x_0, \ldots, x_n)$ of elements of I such that for all $i \geq 1$, $(x_{i-1}, x_i) \in G$ or $(x_i, x_{i-1}) \in G$.

$x_1, \ldots, x_{n-1}$ are called intermediate nodes on this chain.

On an intermediate node x_i a chain can have three connexions as illustrated in Figure 4.

Let (I, G) be a DAG, $S \subseteq I$, a and b be distinct nodes in $\overline{S}$. A chain between a and b is d-separated by S if there is an intermediate node x satisfying one of the two properties:

(i) $x \in S$ and the connection is, on x, serial or diverging,

(ii) $x \notin S^+$ and the connection is converging on x.

In other words, A chain is not d-separated by S if it is in a converging connection at each intermediate node of S, and in a serial or a diverging connection at each intermediate node that has no descendants in S.

Classic Result 1

If A and B are d-separated by S. then the variables X_A and X_B are independent conditional on X_S.

### 4.2. Notions

Given a subset C of I.

> *Markov Blanket.* $F(C)$ of a subset C is made of the parents of C, the children of C and the variables sharing a child with C.
>
> *Markov Boundary* of C. $M(C) = C \cup F(C)$
>
> *Close Descent* of C. $T(C)$ is the set of all close descendants of the elements in C other that those in C itself.
>
> *Exterior Roots* of C. $R(C)$ is the set of the parents of the elements of $C \cup T(C)$ other that those in $C \cup T(C)$ itself.

As we can see on the following example:
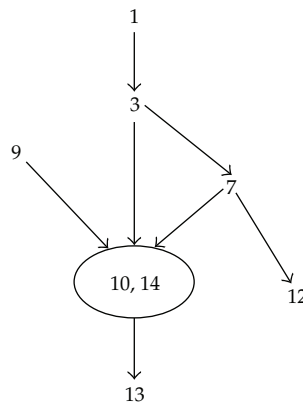
$F(C) = \{1, 10, 14, 12, 9\}$, $M(C) = \{3, 7, 1, 10, 14, 12, 9\}$, $T(3, 7) = \{10, 14, 12\}$, $R(3, 7) = \{1, 9\}$.

Classic Result 2

X_C is d-separated from the rest of the variables conditional on the Markov blanket of C. The proof of these two results and other results related to D-separation can be found in [11].

4.3. Moral and Hypermoral Graphs

Another classic graphic property used with some inference algorithms that we can find in the literature is the notion of the Moral graph.
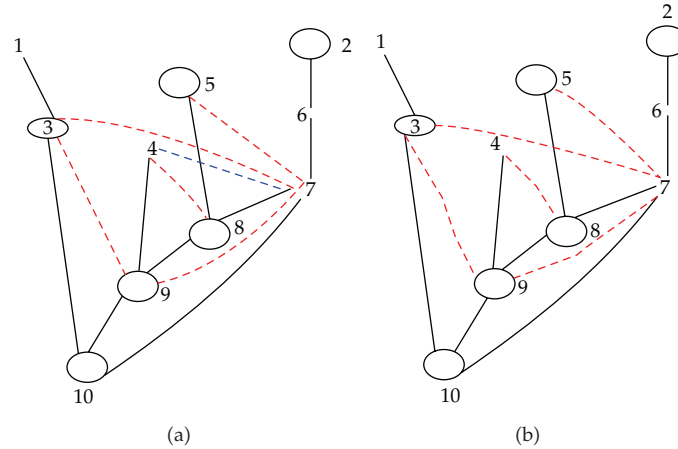
(a)   (b)

Figure 5: Hypermoral and Moral Graphs.

Moral Graph

Given a DAG (I, G)

its associated moral graph is the undirected graph (I, G_m), where G_m is the set

(i) of pairs $\{i, j\}$ such that $(i, j) \in G$ or $(j, i) \in G$,

(ii) of pairs $\{i, j\}$ such that i and j have a child in common.

In a similar way, we define what we call the hypermoral graph defined as follow:

Hypermoral Graph

Given a DAG (I, G)

its associated hypermoral graph is the undirected graph (I, G_{hm}), where G_{hm} is the set:

(i) of pairs $\{i, j\}$ such that $(i, j) \in G$ or $(j, i) \in G$,

(ii) of pairs $\{i, j\}$ such that i and j have a close descendant in common.

In Figure 5, there is a link between 4 and 7 in the hypermoral graph because they share 9 as a close descendant.

4.4. Moral and Hypermoral Partitions

The moral graph helps defining the moral partition as follow.

(i) We call a S-moral partition the partition of $S^+ - S$, denoted $\mathcal{P}^m_{S^+-S}$, defined by the equivalence relation $\sim_{S^+-S}$, where $x \sim_{S^+-S} y$ means " there exists a chain from x to y, in $S^+ - S$, not blocked by S."

In an equivalent way "there exists a chain, in the moral graph G_m, connecting x to y without an intermediate node in S."
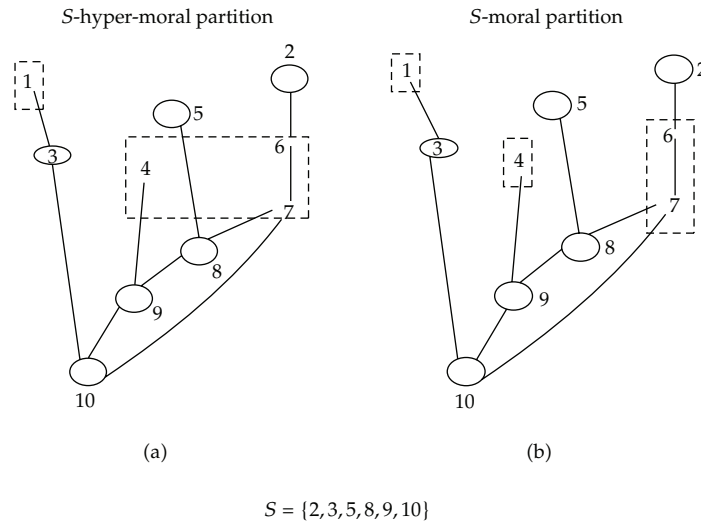
$S = \{2, 3, 5, 8, 9, 10\}$

Figure 6: (a) $p_{\underline{S}}^{\mathrm{hm}}$ The S-hypermoral partition. (b) $p_{\underline{S}}^{m}$ The S-moral partition.

In a similar way we define the hypermoral partition.

(ii) We call an S-hypermoral the partition of $S^+ - S$, denoted $p_{S^+-S}^{\mathrm{hm}}$, defined by the equivalence relation $\approx_{S^+-S}$, where $x \approx_{S^+-S} y$ means "there exists a chain, in the hypermoral graph G_{hm}, connecting x to y without an intermediate node in S". As Illustrated in Figure 6.

4.5. Results

The following results show the possibility of segmenting the computation of the probability distribution P_S.

Theorem 4.1. *Let (I, G, P_I) be a BN, and let S be a subset of I. Let $p_{S^+-S}^{hm}$ be the S-hypermoral partition of $S^+ - S$ and let K be the set of elements of S which are not close descendants of any element of $S^+ - S$, that is, $K = \{k \in S, \ \forall y \in S^+ - S, \ k \notin cd(y)\}$. Then,*

$$P_S = \prod_{k \in K} P_{k/p(k)} \prod_{C \in p_{S^+-S}^{hm}} P_{T(C)/R(C)}, \tag{4.1}$$

where

$$P_{T(C)/R(C)} = \sum_C \prod_{i \in C \cup T(C)} P_{i/p(i)} \left(x_i / x_{p(i)} \right). \tag{4.2}$$

The proof of the theorem can be found in [14].

Theorem 4.2. *The set Q_s of singletons $\{k\}$, where $k \in K$ (if $K \neq \emptyset$), and of subsets $T(C) \cap S^+$, where $C \in p_{S^+-S}^{hm}$, constitutes a partition of S. As Illustrated in Figure 7.*
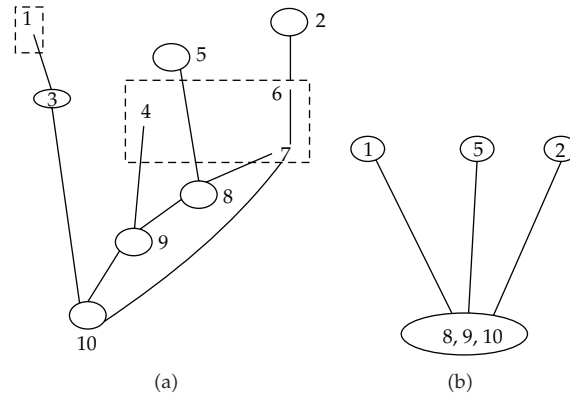
Figure 7: (a) $p_{\overline{S}}^{\mathrm{hm}}$ the S-Hypermoral Partition. (b) BN2 Representing P_S.

5. Unique Partition

We have seen in the last two sections that the application of the SRA for the computation of P_S provides a structure of BN2, and that the use of D-Separations properties allows the segmentation of the computation of P_S and provides also a structure of a level two Bayesian network on S. In fact the two obtained structures are same, this results is giving by the following theorem.

Theorem 5.1. *The following two sets.*

(1) *The subsets $T(C) \cap S^+$, where $C \in p_{S^+ - S}^{hm}$.*

(2) *The set Q_s of singletons $\{k\}$, where $k \in K$ (if $K \neq \emptyset$)*

constitute a unique partition defining a BN2 on S.

Interpretation

This theorem indicates that the level two Bayesian network, characterizing the probability distribution P_S, obtained by application of the SRA, is unique independently of the choices done while running the algorithm. This unique partition is constituted of sets of the two types 1 and 2 mentioned above.

Proof. Let us show that the partition of the target S consists of the subsets of types 1 and 2 as mentioned above in the theorem, in other words consists of $T(C) \cap S^+$ where $C \in p_{S^+ - S}^{\mathrm{hm}}$, and $\{k\}$ for all $k \in K$.

As S^+ is a BN-containing S, without a loss of generality, we can limit ourselves to the case where $S^+ = I$.

The application of the SRA for the computation of P_S requires marginalizing out the set of variables of $\overline{S} = I - S$ following a specific order. Let try to show that the obtained partition by application of the SRA is same mentioned above.

Let us proof this result by induction on the cardinality of $\overline{S}$.

Let us assume that $\mathrm{Card}(\overline{S}) = 1$. In this case $\overline{S}$ has only one element, $\overline{S} = \{i\}$.

On one side, marginalizing out the variable (á) i, by application of the SRA, creates a new node E_i that contains the close descendants of i (i.e., $E_i = \mathrm{cd}(i)$).
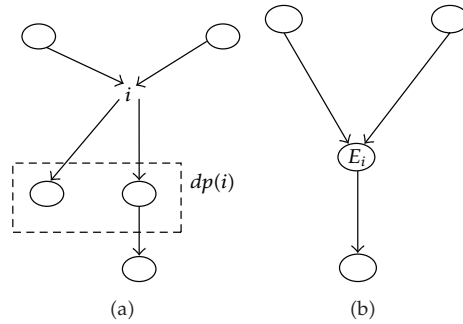
Figure 8: (a): fraction of a BN before marginalizing out i. (b) fraction of a BN2, on S, obtained after summing out i.
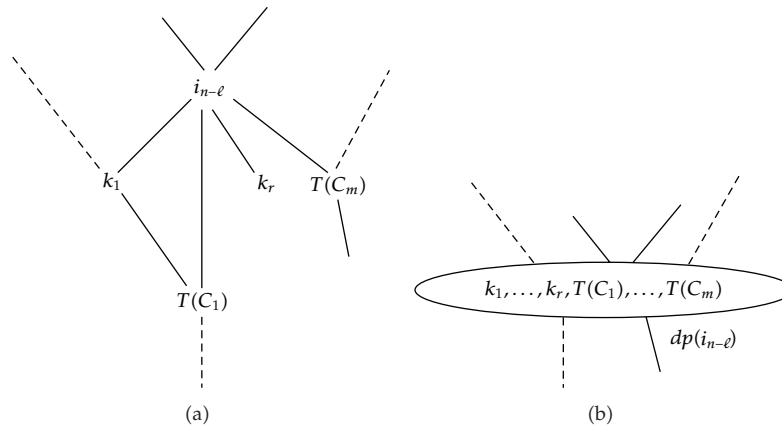

Figure 9: (a): fraction of a BN2 on A. (b): BN2 obtained ater summing out $i_{n-\ell}$.

The BN2 resulting from this marginalization is formed of the new node E_i along with all other remaining nodes, in other words all the $\{k\}$ such that $k \in I - (i_1 \cup \mathrm{cd}(i_1)) = S - E_i$, which is shown in Figure 8.

On the other side, since $\mathrm{Card}(\overline{S}) = 1$, $p_{\overline{S}}^{\mathrm{hm}}$ is constituted of a unique equivalent class, $C = \{i\}$, so, by definition, the partition of S is constituted of

(1) $T(C) \cap S^+ = \mathrm{cd}(i) \cap I = \mathrm{cd}(i)$,

(2) all other nodes, in other words the nodes $k \in K$ such that

$$K = \{k \in S, \ \forall y \in S^+ - S, \ k \notin \mathrm{cd}(y)\} = I - (i \cup \mathrm{cd}(i)) = S - \mathrm{cd}(i) = S - T(C). \qquad (5.1)$$

This shows the result in this first case.

Let us suppose now that $\mathrm{Card}(\overline{S}) = n$, and $(i_1, \ldots, i_n)$ as an hierarchical order on $\overline{S}$.

We are going to sum out following the inverse order of the given hierarchical order.

Let us assume that the result is right till step ℓ (in other words marginalizing out $i_{n-\ell+1}$) and let's proof the result for step $\ell + 1$ (in other words marginalizing out $i_{n-\ell}$).

What justify the proof by induction is that once the marginalizing out $i_n, \ldots, i_{n-\ell+1}$ these last elements will not interfere in the next steps of marginalizing out $i_{n-\ell}, \ldots, i_1$.

The result is right till step ℓ means that we dispose on $A = S \cup \{i_1, \ldots, i_{n-\ell}\}$, of a partition that contains the subsets $\{T(C) \cap S^+\}$ for $C \in \mathcal{P}^{\text{hm}}_{A^+-A}$, and the nodes $\{k\}$ such that $k \in I - \bigcup_{C \in \mathcal{P}^{\text{hm}}_{A^+-A}} (C \cup T(C))$, (i.e., $k \in A - \bigcup_{C \in \mathcal{P}^{\text{hm}}_{A^+-A}} T(C)$), where $\mathcal{P}^{\text{hm}}_{A^+-A}$ is the S-hypermoral partition associated to A.

Showing the result for $\ell + 1$ means to find a partition of $S \cup \{i_1, \ldots, i_{n-\ell-1}\}$ constituted of subsets of type 1 et 2 mentioned above.

On one hand, let's first try to find the partition obtained by application of the SRA. We know that, the marginalization of $i_{n-\ell}$ creates a new node, $E_{i_{n-\ell}}$, which contains the close descendants of $i_{n-\ell}$ in the BN2 on A ($E_{i_{n-\ell}} = \text{cd}(i_{n-\ell})$), shown by Figure 9.

So if we write $J = \{k_1, \ldots, k_r\}$, the set of K such that for all $j \in \{1, \ldots, r\}$, $k_j \cap E_{i_{n-\ell}} \neq \emptyset$ and $L = \{C_1, \ldots, C_m\}$, the set of $\mathcal{P}^{\text{hm}}_{A^+-A}$ such that for all $s \in \{1, \ldots, m\}$, $T(C_s) \cap \text{cd}(i_{n-\ell}) \neq \emptyset$.

In this case, the partition of $B = A - \{i_{n-\ell}\} = S \cup \{i_1, \ldots, i_{n-\ell-1}\}$ is constituted of the new node $E_{i_{n-\ell}} = \text{cd}(i_{n-\ell}) = L \cup J$, and all the other nodes that are left isolated, in other words all the nodes $k \in (K - J)$ and all the $\{T(C) \cap A^+\}$ where $C \in (\mathcal{P}^{\text{hm}}_{A^+-A} - L)$.

If we write $K' = (\mathcal{P}^{\text{hm}}_{A^+-A} - L) \cup (K - J)$ so the partition of B by application of the SRA is composed of the following two types of sets:

(1) $T(C) \cap A^+ = \text{cd}(i_{n-\ell}) = E_{i_{n-\ell}} = L \cup J$,

(2) all other isolated nodes, that is, the nodes of K'.

On the other hand, let's now try to determine the partition of $B = A - \{i_{n-\ell}\}$ using the D-Separation properties.

We have on A a partition composed of nodes $\{T(C) \cap A^+\}$ where $C \in \mathcal{P}^{\text{hm}}_{A^+-A}$, and all nodes $\{k\}$ such that $k \in (A - \bigcup_{C \in \mathcal{P}^{\text{hm}}_{A^+-A}} T(C))$.

Since $B = A - \{i_{n-\ell}\}$, and no descendant of $i_{n-\ell}$ is in $i_1, \ldots, i_{n-\ell-1}$, (by definition of the hierarchical order), so only one equivalent subset is associated, $\mathcal{P}^{\text{hm}}_{B^+-B} = \{\{i_{n-\ell}\}\}$, it results by definition, the partition of $B = S \cup \{i_1, \ldots, i_{n-\ell-1}\}$ is composed of the following types of subsets:

(1) $T(C) \cap B^+ = T(\{i_{n-\ell}\}) = \text{cd}(i_{n-\ell}) = E_{i_{n-\ell}} = L \cup J$,

(2) all other isolated nodes R such that

$$R = \{r \in B, \ \forall s \in B^+ - B, \ r \notin \text{cd}(s)\}. \tag{5.2}$$

So we have $R = B - (i_{n-\ell} \cup \text{cd}(i_{n-\ell})) = A - \text{cd}(i_{n-\ell}) = A - E_{i_{n-\ell}} = A - (L \cup J) = K'$.
This shows that this partition is same partition obtained by application of the SRA. □

References

[1] F. V. Jensen, *An Introduction to Bayesian Networks*, UCL Press, 1999.

[2] F. V. Jensen, S. L. Lauritzen, and K. G. Olesen, "Bayesian updating in causal probabilistic networks by local computations," *Computational Statistics Quarterly*, vol. 5, no. 4, pp. 269–282, 1990.

[3] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, vol. 81 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1993.

[4] G. Dan and J. Pearl, "Axioms and algorithms for inferences involving conditional independence," Tech. Rep. CSD 890031, R-119-I, Cognitive Systems Laboratory, University of California, Los Angeles, Calif, USA, 1989.

[5] D. Heckerman, "A tutorial on learning with Bayesian networks," in *Learning in Graphical Models*, M. Jordan, Ed., MIT Press, Cambridge, Mass, USA, 1999.

[6] N. L. Zhang and D. Poole, "A simple approach to bayesian network computations," in *Proc. of the Tenth Canadian Conference on Artificial Intelligence*, pp. 171–178, 1994.

[7] R. Dechter, "Bucket elimination: a unifying framework for probabilistic inference," in *Uncertainty in Artificial Intelligence*, pp. 211–219, Morgan Kaufmann, San Francisco, Calif, USA, 1996.

[8] L. Smail and J. P. Raoult, "Successive restrictions algorithm in Bayesian networks," in *Proceedings of the International Symposium on Intelligent Data Analysis (IDA '05)*, A. F. Famili et al., Ed., vol. 3646 of *Lecture Notes in Computer Science*, pp. 409–418, Springer, Berlin, Germany, 2005.

[9] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*, Statistics for Engineering and Information Science, Springer, New York, NY, USA, 1999.

[10] R. E. Neapolitan, *Probabilistic Reasoning in Expert Systems*, A Wiley-Interscience Publication, John Wiley & Sons, New York, NY, USA, 1990.

[11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, The Morgan Kaufmann Series in Representation and Reasoning, Morgan Kaufmann, San Francisco, Calif, USA, 1988.

[12] P. Hájek, T. Havránek, and R. Jiroušek, *Uncertain Information Processing in Expert Systems*, CRC Press, Boca Raton, Fla, USA, 1992.

[13] G. Shafer, *Probabilistic Expert Systems*, vol. 67 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, Pa, USA, 1996.

[14] L. Smail, "D-separation and level two Bayesian networks," *Artificial Intelligence Review*, vol. 31, no. 1–4, 2009.